Machine Learning Methods and Technologies for Ubiquitous Computing

Agnese Pinto
Supervisors: Eugenio Di Sciascio, Michele Ruta

DEE, Politecnico di Bari
via Re David 200 – I-70125 Bari, Italy
agnese.pinto@poliba.it

Abstract. The research proposal concerns the use of machine learning techniques for data mining in pervasive environments. It will lead to the formalization of a framework, able to translate series of "raw" data in high-level knowledge. Novel machine learning approaches, interpreting data coming from the environment that surrounds users, will be liveraged. Data will be collected through micro-components deployed in the field and will be processed for the identification and characterization of phenomena and contexts. Eventually they will be semantically annotated to support futher application-level logic-based reasoning and knowledge discovery.

1 Scenario

The Ubiquitous Computing paradigm, introduced by Mark Weiser [16], refers to manifold aspects involving pervasiveness in information storage, processing and discovery. It is aimed to a model of human-computer interaction where information as well as computational capabilities are deeply "integrated" into everyday objects and/or actions. In such vision, the increasing "availability" of processing power would be accompanied by its decreasing "visibility". As opposed to classical paradigms, where a user explicitly engages a single device to perform a specific task, exploiting ubiquitous computing features the user will interact with many computational devices simultaneously. She will extract data from the objects permeating the environment during ordinary activities, even not necessarily being aware of what is happening. By embedding short-range mobile transceivers into a wide array of devices and everyday objects, new kinds of interactions can be enabled between people and things, as well as between things themselves. This is the Internet of Things (IoT) vision [12]. The IoT paradigm is based on the use of a large number of heterogeneous micro devices, each conveying a small amount of useful information. Several emerging technologies are suitable to bridge the gap between physical things and the digital world. For instance, Radio Frequency IDentification (RFID) [7] allows object identification by means of electronic transponders (tags) attached to items. Wireless Sensor Networks (WSNs), on the other hand, allow to monitor environmental parameters, supporting queries and automatic alerts triggered by application-defined events [9]. Both these technologies are characterized by the dissemination of unobtrusive, inexpensive and disposable micro-devices in a given environment.

From the user's standpoint, the goal of pervasive computing is to reduce the amount of user effort and attention required to benefit from computing systems. Current mobile resource discovery protocols have been directly derived from the ones originally designed for infrastructure-based wired networks. In fact, current technologies such as RFID, Bluetooth [2] and ZigBee [18] only allow string matching for item identification. Nevertheless, purely syntactic match mechanisms cannot support more advanced wireless applications, since they provide only boolean yes/no outcomes. It is desirable to manage requests and service descriptions with richer and unambiguous meaning, by adopting formalisms with well-grounded semantics. Wireless communication technologies and mobile computing systems are approaching sufficient maturity to overcome the above limitations. In particular, an advanced mobile resource discovery facility should be able to support non-exact matches [3] and to provide a ranked list of discovered resources or services. This allows to satisfy a user request "to the best possible extent" whenever fully matching resources/services are not available. To do so, techniques for Knowledge Representation (KR) and semantic-based matchmaking may be useful [6]. Hence the goal of pervasive knowledge-based systems is to embed semantically rich and easily accessible information into the physical world.

Nevertheless, in order to cope with pervasive computing contraints, techniques for objects and phenomena characterization are strongly needed. In fact, the context characterization is a foundamental issue in the semantic annotation of a scenario which is mandatory for further dicovery stages.

2 Proposal

Design and impelmentation of effective pervasive computing frameworks involve the study and testing of new technologies. Such technologies should allow to interpret in an automatic or semi-automatic way data extracted from the objects dipped in a generic context, traslating them in knowledge useful to automate processes or support user decisions and activities. In pervasive computing context, data often exhibit a high level complexity (different sensing/collecting technologies, huge volumes, inter-dependency relationships between sources) and dinamicity (real-time update and critical aging). There is the need for methods and algorithms characterized by accuracy, precision and timeliness, also taking into account the limits - in terms of computing resources, memory and communication - of devices.

This research proposal therefore aims to study and formalize methodologies and tools for managing data streams; data collected from a large number of micro-devices in specific environments, using data mining techniques, in particular machine learning algorithms. The semantic-based technologies for element annotation combined with Machine Learning (ML) theory can lead to the formalization of innovative frameworks for pervasive computing, able to transform field data streams in knowledge. Such knowledge will be then usable in application-level decisional processes, to improve the human activities. In other words, raw data will be gathered through a series of micro-components deployed and dipped in a given environments; the collected data will be processed for the identification and characterization of phenomena and contexts. The phenomena and contexts so identified will be described through semantic annotations, in order to be processed by automated logic-based reasoning algorithms.

Data Mining is the process of discovering interesting hidden knowledge by identifying patterns from large amounts of data, where the data can be stored in databases, data warehouses or other information repositories. ML is an Artificial Intelligence area that provides the technical basis of Data Mining. ML manages an abstraction process which happens taking the data and inferring whatever structure underlies them [17], *i.e.*, the main goal of research in machine learning is to "learn" to automatically recognize complex patterns and to make intelligent decisions based on data already analyzed. A peculiarity of ML is *induction*, the extraction of general laws from a set of observed data. It is opposed to *deduction* where, starting from the general laws, the expected value of a set of variables is determined. Induction begins with the observation aiming at measuring a set of variables and then make predictions for further information. This process is named inference.

To carry out the proposed research we will proceed to identify the mainly used methods, with reference to mobile and pervasive computing environments to highlight both strengths and limits. We then will proceed to the definition of innovative annotation methodologies, through the integration and adaptation of existing technologies and/or the definition of new models, algorithms and techniques for machine learning and data analysis. Finally we will apply research outcomes in selected case studies under different scenarios.

3 Applications

Several approaches already exist devoted to pattern recognition to identify and characterize phenomena, events or activities of users in pervasive environments [1][8][11][4]. However, they have significant limitations in the characterization phase and in the support to the discovery of information resources. They are unable to characterize - al least partially - situations that do not correspond exactly to the rules established in the learning and system configuration phase. There are few approaches in literature that support advanced reasoning for mobile devices, which use inference algorithms for discovery, ranking and explanation of retrieval results.

Pervasive computing has many potential applications, from intelligent workplaces and smart homes to healthcare, gaming, leisure systems and to public transportation [10][13][5]. The approach and studies previously sketched could be applied in most of them.

In Wireless Semantic Sensor and Actor Networks (WSSANs) scenarios, using semantic-based techniques at the application layer, it is possible to define

Wireless Semantic Sensor Networks (WSSN) able to offer more versatile services than traditional WSN [14]. For example in a scene of danger, thanks to a set of sensors for the detection of environmental conditions, it is possible to interpret detected data using machine learning techiques and identify one or more events. Recognized events can be semantically annotated and such semantic descriptions can be employed for making inferences.

In infomobility and driving assistance context, for example, several frameworks hare been proposed that aim to improve the capabilities of navigation systems (see as an instance [15]). It provides an approach that could be futher used to retrieve and deliver information to regulate vehicular traffic and to improve the safety of travel. In these scenarios, the contribution of data mining is essential to extract high-level information from a large number of low-level parameters that can be scanned from the car (through the on-board diagnostics OBD protocol - http://www.arb.ca.gov/msprog/obdprog/obdprog.htm) and built-in micro-components of a smartphone device, to accurately characterize the system (vehicle + driver + environment) and to improve driving safety and efficiency.

4 Conclusion

The presented research proposal refers to the extraction of knowledge from pervasive contexts using ML technologis and as Weiser said: "ubiquitous computers will help overcome the problem of information overload. There is more information available at our fingertips during a walk in the woods than in any computer system, yet people find a walk among trees relaxing and computers frustrating. Machines that fit the human environment instead of forcing humans to enter theirs will make using a computer as refreshing as taking a walk in the woods" [16].

References

1. Agarwal, I., Krishnaswamy, S., Gaber, M.M.: Resource-aware ubiquitous data stream querying. In: International Conference on Information and Automation (2005)
2. Bluetooth: http://www.bluetooth.com
3. Chakraborty, D., Perich, F., Avancha, S., Joshi, A.: Dreggie: Semantic service discovery for m-commerce applications. In: Workshop on Reliable and Secure Applications in Mobile Environment, 20th Symposiom on Reliable Distributed Systems (2001)
4. Choi, S., Kim, J., Kwak, D., Angkititrakul, P., Hansen, J.: Analysis and classification of driver behavior using in-vehicle can-bus information. In: Biennial Workshop on DSP for In-Vehicle and Mobile Systems. pp. 17–19. Citeseer (2007)
5. Chon, J., Cha, H.: Lifemap: A smartphone-based context provider for location-based services. Pervasive Computing, IEEE 10(2), 58–67 (2011)
6. Colucci, S., Di Noia, T., Pinto, A., Ruta, M., Ragone, A., Tinelli, E.: A nonmonotonic approach to semantic matchmaking and request refinement in e-marketplaces. International Journal of Electronic Commerce 12(2), 127–154 (2007)

7. Epcglobal Inc.: EPC Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocolfor Communications at 860 MHz–960 MHz. Tech. rep., EPC-global (January 2005)
8. Fan, W., an Huang, Y., Wang, H., Yu, P.S.: Active mining of data streams. In: Fourth SIAM International Conference on Data Mining. pp. 457–461 (2004)
9. Gracanin, D., Eltoweissy, M., Wadaa, A., DaSilva, L.A.: A service-centric model for wireless sensor networks. IEEE Journal on Selected Areas in Communications 23(6), 1159–1166 (2005)
10. Hagras, H., Callaghan, V., Colley, M., Clarke, G., Pounds-Cornish, A., Duman, H.: Creating an ambient-intelligence environment using embedded agents. IEEE Intelligent Systems 19(6), 12–20 (Nov 2004), http://dx.doi.org/10.1109/MIS.2004.61
11. Hina, M., Tadj, C., Ramdane-Cherif, A.: Machine learning-assisted device selection in a context-sensitive ubiquitous multimodal multimedia computing system. In: Industrial Electronics, 2006 IEEE International Symposium on. vol. 4, pp. 3014–3019. IEEE (2006)
12. ITU: Internet Reports 2005: The Internet of Things. Tech. rep., ITU (November 2005)
13. Noguchi, K., Somwong, P., Matsubara, T., Nakauchi, Y.: Human intention detection and activity support system for ubiquitous autonomy. In: Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on. vol. 2, pp. 906–911. IEEE (2003)
14. Ruta, M., Scioscia, F., Di Noia, T., Di Sciascio, E.: A hybrid zigbee/bluetooth approach to mobile semantic grids. Computer Systems Science and Engineering 25(3), 235 (2010)
15. Ruta, M., Scioscia, F., Ieva, S., Di Sciascio, E.: T$\hat{2}$.o.m. t.o.m.: Techniques and technologies for an ontology-based mobility tool with open maps. Current Trends in Web Engineering pp. 199–210 (2010)
16. Weiser, M.: The computer for the 21st century. ACM SIGMOBILE Mobile Computing and Communications Review 3(3), 3–11 (1999)
17. Witten, I.H., Frank, E., Hall, M.A.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Amsterdam, 3. edn. (2011)
18. ZigBee Alliance: ZigBee Specification. ZigBee Alliance, 053474r17 edn. (January 17 2008)